人工知能学会研究会資料
SIG-FPAI-B404-08

距離ベースの特徴選択指標

A Theory of Feature Selection Measures

申　吉浩[1]*　Adrian Pino Angulo[2]　久保山哲二[3]

[1] 兵庫県立大学　[2] University of Holguin　[3] 学習院大学

Abstract: We present a method to define feature selection measures based on metrics. Also, we show that the well-known Bayesian risk can be derived from some metric and give a new characterization to it.

1　距離ベースの特徴選択指標

テータセットの確率分布を p-norm による数列空間の要素と見なすことで、データセットと特徴集合のペアを距離空間に埋め込むことが可能である。この埋め込みにより、以下のように、特徴選択指標を定義することが可能である。

For $1 \leq p < \infty$,

$$\widehat{\mu_{\ell^p}}(\boldsymbol{p}) = \sqrt[p]{\sum_{i\in\mathbb{N}} \left(\sum_{j\in\mathbb{N}} \boldsymbol{p}(i,j)^p - \boldsymbol{p}(i,\overline{j}(i))^p \right)};$$

For $p = \infty$,

$$\widehat{\mu_{\ell^\infty}}(\boldsymbol{p}) = \max\left\{ \boldsymbol{p}(i,j) \,|\, (i,j)\in\mathbb{N}^2, (i,j)\neq(i,\overline{j}(i)) \right\}.$$

$p = 1$ の時、この指標はよく知られたベイズリスクと一致し、$p > 1$ の時は今迄に知られていない新規の指標となる。この論文では、これらの指標と、実際の分類器による分類精度との関連を調べる。

2　指標 $\widehat{\mu_{\ell^p}}$ の値と分類精度の相関

よい指標は分類精度と強い相関をもっているであろうと期待することは自然である。この節では、$\widehat{\mu_{\ell^p}}$ $(p = 1, 2, \ldots, 5)$ について、分類精度との相関を調べる。特に、$p = 1$ と $p > 1$ の場合の比較に焦点を当てる。$\widehat{\mu_{\ell^1}}$ はよく知られているベイズリスクであり、$p > 1$ の場合の $\widehat{\mu_{\ell^p}}$ は今迄に知られていない新規の指標である。

2.1　データセット

表 1 に、実験で用いたデータセットと、その主な属性（名称、特徴数、サンプル数）を記す。

2.2　実験手順

実験手順を述べる

2.2.1　特徴集合のサンプリング

表 1 で示した各データセットについて、60 個の特徴集合をランダムに選択する。特徴集合の大きさもランダムに変動する。全てのデータセットにわたる総計で、$60 \times 12 = 720$ 対の特徴集合とデータセットのペアが得られる。

Table 1: Datasets

Name	#Feat.	#Exam.	#Classes
Arrhythmia	279	452	13
Audiology	69	226	24
Mfeat-Factor	216	2000	10
Mfeat-Fourier	76	2000	10
Mfeat-Karhunen	64	2000	10
Mfeat-Pixel	240	2000	10
Mfeat-Zernike	47	2000	10
Musk	166	476	2
Optidigits	64	5620	10
Sonar	60	208	2
Spambase	57	4601	2
Spectrometer	100	531	48

2.2.2　データセットの局所化

特徴集合とデータセットの各対に対して、その特徴集合に含まれない全ての特徴を除去することにより、データセットの局所化を行う。結果として、720 個の局所化されたデータセットを得る。

2.2.3　分類精度の計測

Naïve Bayes、C4.5 及び SVM の三種類の分類器を、得られた局所化データセットのそれぞれに適用し、10-フォルド交叉検証により得られる AUC-ROC の値の平均値を、分類精度として記録する。

2.3　Results

図 1、2 及び 3 は、それぞれ、Naïve Bayes、C4.5 及び SVM による実験結果の散布図を示す。x 軸は、(a) $\widehat{\mu_{\ell^1}}$、(b) $\widehat{\mu_{\ell^2}}$、(c) $\widehat{\mu_{\ell^3}}$、(d) $\widehat{\mu_{\ell^4}}$ 及び (e) $\widehat{\mu_{\ell^5}}$ による指標値であり、y 軸は AUC-ROC の平均値を表す。図から、$p > 1$ の時の $\widehat{\mu_{\ell^p}}$ の値は、$\widehat{\mu_{\ell^1}}$ に比較して、強い負の相関を示すことが観察できる。

*yshin@ai.u-hyogo.ac.jp

2.4 分析

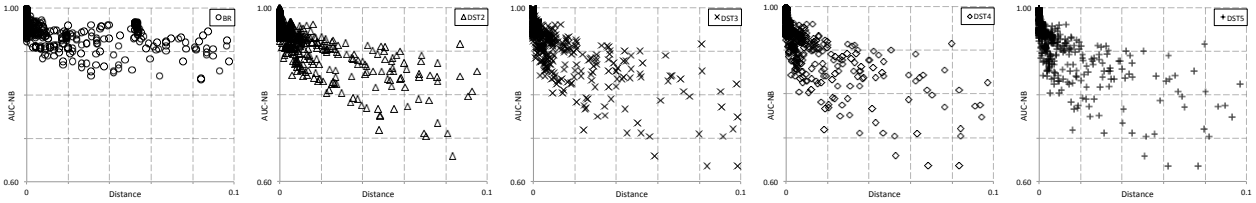
$\widehat{\mu_{\ell^1}}$ と $\widehat{\mu_{\ell^p}}$ $(p>1)$ とを、分類精度との相関の観点から比較するために、関数 $P_t(x)$ を導入する。N_x を $\widehat{\mu_{\ell^p}}$ 距離が $[x, x+0.01)$ の間にあるプロットの数とし、$N_{t,x}$ を AUC-ROC の値が t を超え、かつ、N_x を $\widehat{\mu_{\ell^p}}$ 距離が $[x, x+0.01)$ の間にあるプロットの数とし、$P_t(x)$ を $P_t(x) = \frac{N_{t,x}}{N_x}$ により定義する。直感的には、$P_t(x)$ は、指標値が x の値を取ったとき、分類精度が t を超える確率の近似となる。

図 4、5 及び and 6 で、$P_t(x)$ の曲線を、Naïve Bayes、C4.5 及び SVM について図示する。t の値は $\{0.95, 0.90, 0.85, 0.80\}$ から選ぶものとする。$P_t(x) \geq P_{t'}(x)$ が $t < t'$ に対して常に成り立つので、t に対する曲線は t' に対する曲線の上に来る。図より、以下の性質が読み取れる。

- $p > 1$ の時、$\widehat{\mu_{\ell^p}}$ に対する曲線は類似している。一方、$\widehat{\mu_{\ell^1}}$ の曲線は他の曲線とは著しく形状が異なる。
- $p > 1$ に対する $\widehat{\mu_{\ell^p}}$ 曲線は、$\widehat{\mu_{\ell^1}}$ の場合に比較して、指標値と分類精度の間の相関をより明確に示している。

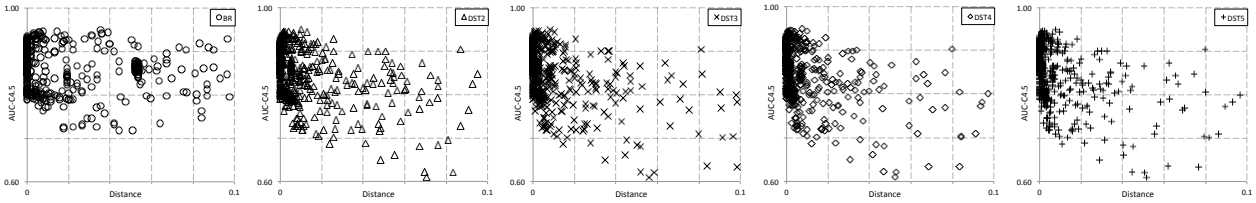
実際、相関係数を計算すると以下の表のようになる。

$t=$	0.95	0.90	0.85	0.80
		Naïve Bayes		
$\widehat{\mu_{\ell^1}}$	−0.42	−0.51	−0.60	−0.52
$\widehat{\mu_{\ell^2}}$	−0.58	−0.79	−0.97	−0.85
$\widehat{\mu_{\ell^3}}$	−0.56	−0.69	−0.76	−0.79
$\widehat{\mu_{\ell^4}}$	−0.54	−0.71	−0.89	−0.89
$\widehat{\mu_{\ell^5}}$	−0.54	−0.73	−0.91	−0.89
		C4.5		
$\widehat{\mu_{\ell^1}}$	−0.17	−0.41	−0.34	−0.38
$\widehat{\mu_{\ell^2}}$	−0.41	−0.47	−0.79	−0.83
$\widehat{\mu_{\ell^3}}$	−0.52	−0.29	−0.72	−0.69
$\widehat{\mu_{\ell^4}}$	−0.52	−0.38	−0.77	−0.81
$\widehat{\mu_{\ell^5}}$	−0.52	−0.44	−0.82	−0.84
		SVM		
$\widehat{\mu_{\ell^1}}$	−0.52	−0.62	−0.55	−0.23
$\widehat{\mu_{\ell^2}}$	−0.52	−0.54	−0.65	−0.80
$\widehat{\mu_{\ell^4}}$	−0.55	−0.52	−0.63	−0.75
$\widehat{\mu_{\ell^4}}$	−0.52	−0.52	−0.59	−0.72
$\widehat{\mu_{\ell^5}}$	−0.52	−0.52	−0.56	−0.71



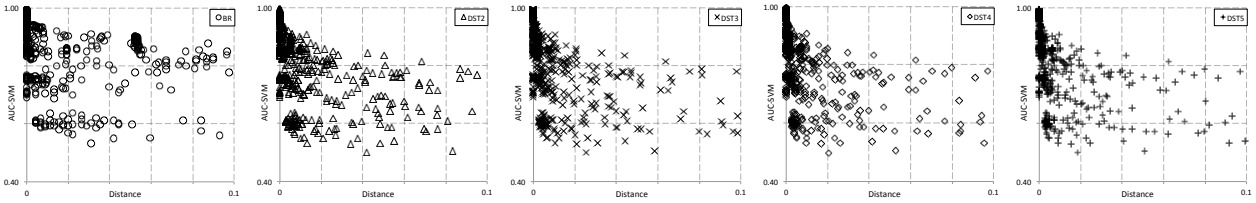
(a) $\widehat{\mu_{\ell^1}}$ (b) $\widehat{\mu_{\ell^2}}$ (c) $\widehat{\mu_{\ell^3}}$ (d) $\widehat{\mu_{\ell^4}}$ (e) $\widehat{\mu_{\ell^5}}$

Figure 1: Scatter plots of the experimental results (Naïve Bayes)



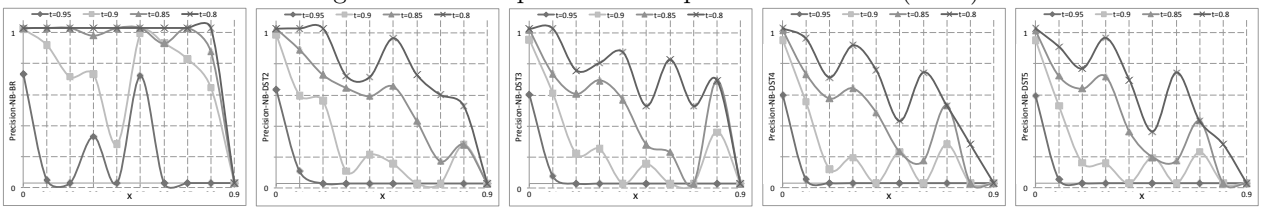
(a) $\widehat{\mu_{\ell^1}}$ (b) $\widehat{\mu_{\ell^2}}$ (c) $\widehat{\mu_{\ell^3}}$ (d) $\widehat{\mu_{\ell^4}}$ (e) $\widehat{\mu_{\ell^5}}$

Figure 2: Scatter plots of the experimental results (C4.5)



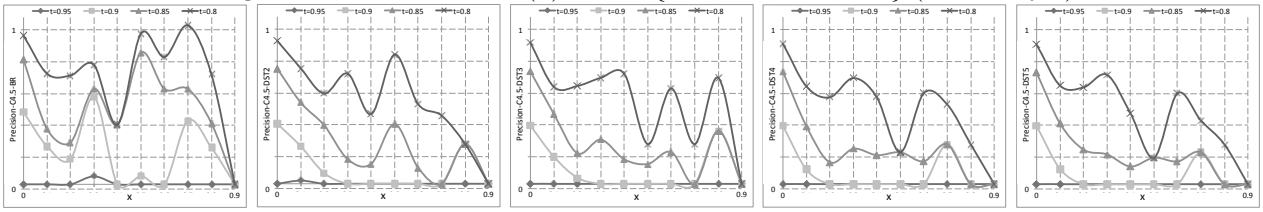
(a) $\widehat{\mu_{\ell^1}}$ (b) $\widehat{\mu_{\ell^2}}$ (c) $\widehat{\mu_{\ell^3}}$ (d) $\widehat{\mu_{\ell^4}}$ (e) $\widehat{\mu_{\ell^5}}$

Figure 3: Scatter plots of the experimental results (SVM)



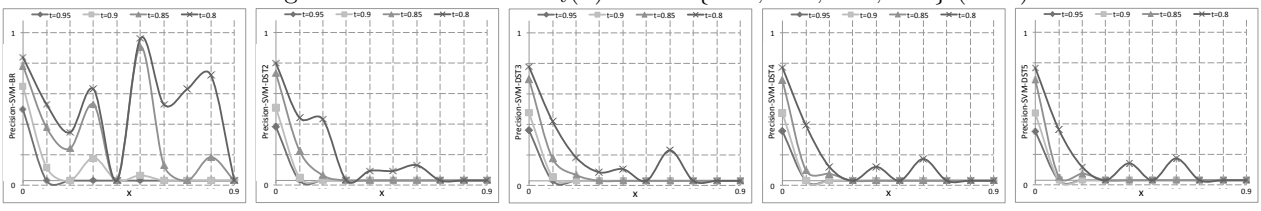
(a) $\widehat{\mu_{\ell^1}}$ (b) $\widehat{\mu_{\ell^2}}$ (c) $\widehat{\mu_{\ell^3}}$ (d) $\widehat{\mu_{\ell^4}}$ (e) $\widehat{\mu_{\ell^5}}$

Figure 4: The curves of $P_t(x)$ for $t \in \{0.95, 0.90, 0.85, 0.80\}$ (Naïve Bayes)



(a) $\widehat{\mu_{\ell^1}}$ (b) $\widehat{\mu_{\ell^2}}$ (c) $\widehat{\mu_{\ell^3}}$ (d) $\widehat{\mu_{\ell^4}}$ (e) $\widehat{\mu_{\ell^5}}$

Figure 5: The curves of $P_t(x)$ for $t \in \{0.95, 0.90, 0.85, 0.80\}$ (C4.5)



(a) $\widehat{\mu_{\ell^1}}$ (b) $\widehat{\mu_{\ell^2}}$ (c) $\widehat{\mu_{\ell^3}}$ (d) $\widehat{\mu_{\ell^4}}$ (e) $\widehat{\mu_{\ell^5}}$

Figure 6: The curves of $P_t(x)$ for $t \in \{0.95, 0.90, 0.85, 0.80\}$ (SVM)